Global Journal of Computer Sciences: Theory and Research

Volume 8, Issue 3, (2018) 111-119

www.gjcs.eu

Graph clustering based size varying rules for lifelong topic modelling

Muhammad Taimoor Khan, Department of Computer Science, Bahria University, Islamabad 44000, Pakistan
Shehzad Khalid*, Department of Computer Engineering, Bahria University, Islamabad 44000, Pakistan
Furqan Aziz, Department of Computer Science IMSciences Peshawar, Peshawar 250000, Pakistan

Suggested Citation:
Khan, M. T., Khalid, S. & Aziz, F. (2018). Graph clustering based size varying rules for lifelong topic modelling. *Global Journal of Computer Sciences: Theory and Research*. *8*(3), 111–119.

Received from March 17, 2018; revised from July 18, 2018; accepted from November 21, 2018.
Selection and peer review under responsibility of Prof. Dr. Dogan İbrahim, Near East University, Cyprus.

Abstract

Lifelong learning topic models identify the hidden concepts discussed in the collection of documents. Lifelong learning models have an automatic learning mechanism. In the learning process, the model gets more knowledgeable with experience as it learns from the past in the form of rules. It carries rules to the future and utilises them when a similar scenario arise in the future. The existing lifelong learning topic models heavily rely on statistical measures to learn rules that lead to two limitations. In this research work, we introduce complex networks analysis for learning rules. The rules are obtained through hierarchical clustering of the complex network that has different number of words within a rule and has directed orientation. The proposed approach improves the utilisation of rules for improved quality of topics at higher performance with unidirectional rules on the standard lifelong learning dataset.

Keywords: networks, lifelong, models, networks analysis

* ADDRESS FOR CORRESPONDENCE**: Shehzad Khalid,** Department of Computer Engineering, Bahria University, Islamabad 44000, Pakistan.
E-mail address: shehzad_khalid@hotmail.com / Tel.: +0092-51-9261281

1. Introduction

Topic models extract latent topics structures from the observable words and documents (Blei, & Jordan, 2006; Blei, Ng & Jordan, 2003). When topic models are applied to a large collection of documents, it analyse the documents and represent them as few conceivable topics (Khan, Durrani, Khan, Ali & Khalid, 2015). Each topic consists of a list of ordered words having high relevance to convey the underlying concept held in the topic (Pang & Lee, 2008). Topic models are predominantly used in natural language processing; however, its applications have outgrown to different research areas (Khan & Khalid, 2016; Khan, Durrani, Ali, Inayat, Khalid & Khan, 2016). As a dimensionality reduction technique, it is also used to extract features for classifiers. As a generative model, the topic model can also produce new documents using a given topic structure. In text analysis, topic models are usually applied to commercial product reviews for identifying product aspects automatically. These aspects are the key features of a product gaining more popularity and are used as opinion targets in Aspect-based sentiment analysis (Khan et al., 2015). For example, the *laptop* reviews may have topics representing *screen, price, portability* and *battery life*. Topic models cater to different applications to explore, categorise, summarise and analyse textual data (Pang & Lee, 2008).

There are different extensions of topic models that use some level of external guidance to improve accuracy at the cost of compromising applicability (Shen, Wu & Li, 2016). Therefore, these approaches cannot be used in building applications that can scale up to the overwhelming flow of data from online applications (Information Resources Management Association, 2016). Lifelong learning approach is specifically designed for such a scenario (Ruvolo & Eaton, 2013; Xu, Yang, Harenberg & Samatova, 2017). It has an automatic learning mechanism that allows the model to learn without external support (Khan & Khalid, 2017). In the process, the model gets more knowledgeable with experience and is reflected in the quality of results (Khan, Azam, Khalid & Yao, 2017). Lifelong learning approach is used with supervised, semi-supervised and transfer learning models as well (Andrzejewski & Zhu, 2009). It makes a very realistic assumption that everything is related to something and need not be learnt every time (Khan & Khalid, 2016). This is in contrary to the traditional machine learning that throws what is learnt from one task and learns about each task from the scratch (Khan, Yar, Khalid & Aziz, 2016). It is similar to transfer learning but learns from all tasks processed in the past and is blind to future tasks. Therefore, the model learns without knowing what may be required in the future and when will it be required (Chen & Liu, 2014a; Khan & Khalid, 2016). However, it keeps on learning what it feels to be of value. Lifelong learning is based on two assumptions to learn continuously (Ruvolo & Eaton, 2013). These assumptions are very realistic in the real world. First, if many people are discussing something that it is important to learn about. Second, if most people are saying the same thing about it, then it is most likely true (Khan et al., 2016). The model learns rules and retains them as knowledge with the impression that a similar scenario may appear in future where it may help to make an informed decision.

The existing lifelong learning topic models heavily rely on statistical measures for learning rules. An exhaustive list of candidate rules is generated as candidate rules (Chen, Mukherjee & Liu, 2014; Khan et al., 2016). The candidate rules undergo an evaluation mechanism to assign a fitness score to each rule (Chen & Liu, 2014a). It is usually based on distance measure between the words of a rule to identify their correlation (Khan & Khalid, 2016). The words that are together more often than that are not form positive candidate rule while the words that are more separate than that are together form negative candidate rules. This leads to having four types of candidate rules i.e., strong positive rules, weak positive rules, weak negative rules and strong negative rules (Khan & Khalid, 2017). We are generally interested in strong rules as they have high support from the underlying data.

Weak rules incur extra computational cost while do not bring notable change in results. At worst case, they may falsely promote unrealistic associations and may cause the results to deteriorate. Two thresholds are applied to the evaluation score of each rule as T_- and T_+. The ones having a positive correlation above T_+ are preserved as positive rules while the ones having negative correlation below

T_{-} are preserved as negative rules. They are also called must-links and cannot-links, respectively (Khan & Khalid, 2017).

There are two limitations with the existing approaches related to learning rules. First, the candidate rules are evaluated only for a predetermined number of words (Chen & Liu, 2014a; Khan & Khalid, 2016). Therefore, the rules are of two words only as their mutual correlation drops considerably as the number of words increases. It leads to a very high number of small rules that result in a high computational cost. Second, multiple small rules within a topic are ineffective towards chain incoherence (Leenders & McCallum, 2011; Mimno, Wallach, Talley). It is a type of topic incoherence where the words make sense in pairs forming a chain but do not present a concise concept collectively as a topic. Third, they do not decide the orientation of the rule and are arbitrary when it comes to transferring the impact of rules (Khan & Khalid, 2016). In such case, both words of a rule attempt to pull the other word towards to topic it resides in. Thus, both words are attempting to influence the other word when they are sampled in their respective turns. The collective objective is to place two words together within a topic to represent it. However, with an arbitrary orientation, the extra computational cost is incurred due to multiple repeated patterns. Although the words of the positive rules settle to be within the sample topic and in different topics for negative rules, the computational cost can be reduced considerably for the intermediate iterations.

2. Literature review

Topic models explore a collection of documents in a unknown domain to extract topics. A topic consists of a group of words having strong contextual relevance (Blei & Jordan, 2006). Although they ignore the word semantics, topic models perform better then frequency-relation based techniques, particularly in specialised domains (Blei, Ng & Jordan, 2003). By varying the number of topics in an analysis, the coarser or finer level topics can be extracted (Khan et al., 2017). Fewer topics focus on global concepts only, while more number of topics have each topic specific to a local concept. As an unsupervised approach, the topics are unlabeled. However, the highest probability words in a topic are chosen to represent it (Khan, Durrani, Khalid & Aziz, 2016). Basically, all words belong to all topics but the lower probability words are ignored. Moreover, a word having a lower probability for one topic may have a higher probability for another topic. The inference technique is responsible for arranging the words into their respective topics after random initialization (Shen et al., 2016). It makes use of word-topic distribution and topic-document distribution to choose the most probable topic for the sampled instance of a word. The most probable topic depends on the topics other words of the same document have and topics of the other instances of the same word. Both the values are normalised with the number of topics in documents and words in topics (Khan & Khalid, 2017). When a word switches to its most probable topic, according to the current state of the model, the two distributions are updated accordingly, i.e., the probability of the sampled word drops for the topic to which it previously belonged and is increased for the new topic (Khanet al., 2016). Similarly, the probability of the previous topics is decreased and the probability of the new topic is increased in the document of the sampled word. The process stops when the words do not switch topics anymore. It represents the final arrangement of words into their respective topics (Blei, Ng & Jordan, 2003; Ramage, Hall, Nallapati & Manning, 2009).

Topic models require a large number of documents to produce compact and coherent topics. It may produce incoherent topics if the dataset have few samples or noise (Khan et al., 2015). Topic incoherence refers to the topics having words that do not hold well together to represent a concept (Andrzejewski & Zhu, 2009). It is evaluated as topic coherence or perplexity as intrinsic evaluation measures (Bouma, 2009). Extrinsic evaluation techniques make use of an external dictionary or resource to evaluate the quality of topics. The types of incoherence are intruded, chained, unbalanced and random topics. Intruded topic incoherence has few words that do not go well together with the other words of the same topic. Chained incoherence has words related in pairs but the overall topics have a mix of local and global terms that are hard to comprehend as a concept (Mimno et al., 2011).

Random topics have words that have no contextual relevance. It may represent noise in the dataset. Topic models are extended to avoid topic incoherence by incorporating labeled training or expert's guidance. They follow supervised, hybrid, semi-supervised, transfer learning and knowledge-based approaches (Khan et al., 2016). Knowledge-based topic models follow a semi-supervised approach but the domain expert provides rules instead of seed tagging (Khan et al., 2017). The rules can be positive or negative where positive rule words stay close to each other within a topic and vice versa (Khan & Khalid, 2017). Although the topic model extensions improved its accuracy, its scalability is compromised as these techniques cannot be applied to a fresh domain for which no labeled training data or expert's intuition is available (Shen et al., 2016). On the other hand, online social data are dynamic and evolving with fresh topics frequently introduced (Khan et al., 2016).

Lifelong learning topic models are used for various tasks in the works of (Xu et al., 2017). The model has a continuous learning mechanism and is also known as Never-ending learning or Human-like learning models. Using a lifelong learning approach with topic models is to enable them to learn their rules automatically. It makes use of the past experience and learns from it, which is utilised in future tasks. Automatic knowledge learning and lifelong topic modelling (LTM) are the initial lifelong learning approaches with topic models (Chen & Liu, 2014a; Chen & Liu, 2014b). They used frequent pattern mining on topic clusters to extract must-link rules. The rules are transferred using generalised Polya urn model (GPU) (Mahmoud, 2008). Automatic must-link cannot-link (AMC) used both must-link and cannot-link rules having high point-wise mutual information (PMI) (Chen & Liu, 2016). They transferred the impact of rules into the inference technique using Multi-GPU (Chen, Mukherjee & Liu, 2014). In the OAMC model, the rules are extracted for sequential flow of domains using normalised PMI technique, while using GPU for transferring their impact (Mimno et al., 2011). A filtering mechanism was also introduced to refine rules and avoid their contradictions (Khan et al., 2016). The rules maintaining higher confidence over the following tasks and have higher utility are retained in the knowledge-base. The model was specifically designed for large-scale data which could experience performance bottleneck at high experience (Khan et al., 2016). The existing lifelong learning topic models have only used statistical approaches to extract rules; however, new ideas from different domains can improve the quality and quantity of rules. Similarly, new ideas are required to extend the contribution of rules from words arrangement into topics arrangement.

Complex networks are graphs that are often very large and very complex and they arise as a result of modelling a real-world problem through graphs (Han, Escolano, Hancock & Wilson, 2012). The fact that graphs can be considered as one of the most generic data structures makes it a unique choice for the analysis of problems in many areas including pattern recognition and machine learning (Mowshowitz & Dehmer, 2012). One of the challenging problems in graph theory is estimating the complexity of a network. The aim here is to characterise the degree of regularity of a complex network. Entropy-based methods have been proved to be very useful in characterising the structure of a complex network (Mowshowitz & Dehmer, 2012). To this end, the Von Neumann entropy of a graph can be computed from the eigenvalues of the normalised Laplacian of a graph (Han et al., 2012). It is shown that the proposed measure tends to be larger in relation to the number of connected components, long paths, and nontrivial symmetries (Passerini & Severini, 2008). A general framework is used to define the entropy of a graph that associates a probability value to each vertex by employing functions that capture structural information of a graph (Aziz, Hancock & Wilson, 2016). It is further extended to the structure that captures the correlation using walks and cycle structures. One of the most widely used applications of complex networks is to gauge the flow of information across a network (Sadilek, Kautz & Silenzio, 2012). For instance, complex networks can be utilised in epidemiology, where the goal is to study and predict the spread of disease in a social network (Peng, Yang, Cao, Yu & Xie, 2017). The graph theory is exploited to construct a social relationship graph to study social influences in a mobile social network (Massoulie, 2014).

3. Proposed methodology

We propose a complex network-based learning module for lifelong topic models. As a lifelong topic model, it learns continuously to have topics representing aspects with the help of the rules available in the knowledge-base. However, unlike the existing lifelong topic models that have used statistical techniques for rules extraction, the proposed approach extract rules through complex network analysis. It is the first attempt to use LTM with a complex network analysis. The objective is to attain rules that are more flexible in terms of their size, depending upon the arrangement of words within the documents. Furthermore, communities as rules are rich in information and indicate community heads as well that decides the orientation of the community as a rule.

Algorithm 1. Complex Network based Lifelong Topic Model

Input: Large-scale data $\{D = D_1, D_2, ..., D_t, ...D_{t+i}\}$ **Output:** Topics
1: Construct complex network
2: Extract communities
3: Evaluate communities
4: Abstract communities as rules and preserving them
5: **while** domain D_{t+i} **do**
6: Sample relevant communities
7: Initialise the model by random word-topic assignment
8: Apply inference technique with bias from rules
9: generate topics T_{t+i}
10: **end while**

Working of the proposed model is shown in **Algorithm 1**. It consumes large-scale data having *t* domains i.e., $D = D_1, D_2, ..., D_t$ as input in order to generate a complex network. The nodes represent the unique words of all the domains *D*. The two nodes share an edge when they co-occur within a document. The weight of the edge increases the co-occurrence of the two words in multiple documents, across all domains. In line 2, hierarchical clustering is applied on the complex network. In graph theory, a Laplacian matrix gives the representation of a matrix to a graph. The Laplacian matrix can then be used to apply different mathematical evaluations for analyzing various properties of the graph. Normalised laplacian is calculated as:

$$L^{sym} = D^{-1/2} L D^{1/2} = I - D^{-1/2} A D^{-1/2} \quad (1)$$

Where *I* is the identity matrix, A is adjacency matrix and D is diagonal matrix. The eigenvalue decomposition of the Laplacian matrix gives us the eigenvalues and the eigenvectors.

The second smallest eigenvalue of the Laplacian is called the algebraic connectivity of a graph, while the eigenvector corresponding to the second smallest eigenvalue is called Fiedler vector. The sign of the Fiedler eigenvector can be used to partition the graph into two sets of vertices, i.e., It indicates a possible split where the nodes with positive value can form one community while the nodes with negative values represent another community. This technique was originally proposed by Fiedler, and it has found applications in many areas. The process stops when a clustering cannot be separated into smaller clusters. The stopping criteria are based on the minimum number of nodes in a cluster or the cost of separating the nodes of a cluster. All the leaf node communities are potential rules, while the higher level communities are ignored. The cost of separating a cluster *comm* into sub-clusters $comm_a$ and $comm_b$ is calculated as,

$$ttCW\ (comm_a,\ comm_b) = \frac{weight\left(e_{w \in comm_a, w^t \in comm_b}\right)}{TotalEdgeWeight} \quad (2)$$

Where the sum of weights of edges between nodes of $comm_a$ and $comm_b$ is divided by the sum of weights of all edges in comm. If the split is affordable, divide *comm* into $comm_a$ and $comm_b$. If the split cost is too high, the community is not divided and is passed on to the evaluation process.

Line 3 evaluate all the communities to associate a fitness score so that only high-quality communities can be selected as rules. The fitness score is based on the compactness of nodes within a community. The communities as knowledge are more flexible in terms of their sizes and structures as they represent the natural arrangement of words. It does not force the words to form rules of fixed sizes. Instead, a rule can be of any number of words as long as they are strongly interconnected. However, not all communities can be used as a source of knowledge. The inter-connectivity of some communities may not be strong enough to trust them as reliable rules. The communities are evaluated with different measures to identify their measure of fitness and communities that do not possess such fitness are filtered out. The communities are evaluated using weighted graph connectivity index. It measures the strength of a community by taking into account the edge weights and the degree distribution of the network and is computed as,

$$W\ ttCI = \sum_{e_{w,w^t}} \frac{weight\left(e_{w,w^t}\right)}{\sqrt{\left(d_w \times d_{w^t}\right)}} \tag{3}$$

where $e_{w,wt}$ represents an edge in the community. The weights of all edges in the community are added up and normalised by dividing with the square root of the product of the degree of two nodes d_w and d_{wt} that are sharing the edge. It is aggregated across all edges of the community to give a measure of associations among the nodes of a community. The communities with lower *W ttCI* are filtered as they may result in low confidence rules. It addresses the first problem related to varying size rules extraction.

In line 4, the selected communities are transformed into rules so that they can be stored and acquired effectively. The community heads are identified to strict the transfer of rule impact from one direction only. Thus, a word belonging to communities does not need to have its impact transferred to all the other words in the community. Instead, all the words in the community are made to follow their heads that set the tone of transferring the impact of rules. Community heads are the nodes with the highest sum of weighted incident edges. The rear follows the topic of their heads depending upon the impact of community. It addresses the second problem related to the arbitrary orientation of the rules into the topics of the given domain.

In lines 5–10, a new domain D_{t+1} is processed with the help of the knowledge learnt from the past D_t domains. The knowledge base is consulted for relevant knowledge communities, in line 6. The rules learnt are preserved in the knowledge-base as a repository. The new domain may or may not be from the domains used at the learning phase. However, with a higher number of domains in the learning phase, it is assumed that the model has relevant rules for the new domain as well. In Line 7, the model is initialised by assigning topics to words at random. In line 8, Collapsed Gibbs Sampling is used as an inference technique to approximate the topic structure that has generated these documents. It randomly samples a word *w* to update its word-topic assignment. If *w* belongs to a relevant rule as community, these are pushed towards the prominent nodes of the community that are labeled as community heads. Otherwise, *w* is assigned a new topic according to the current state of the model. The inference technique stops when it reaches the best arrangement of words into topics with the help of knowledge communities. The involvement of knowledge communities is observed in individual topics. The final topics are generated in line 10. The rules have helped in the better arrangement of words within topics by employing the intuition from the large-scale data i.e., D_t domains that were processed before. It allows the model to misleading associations among words in case of limited documents or noise as the past rules may not agree to them.

Table 1. Selected communities learnt from 50 tasks

Comm Sr.No.	Size	Avg. Node Degree	Top Words
1	7	56.893	Adaptor, cam, filter, stand, indicator
2	13	12.153	Exception, weird, superior, break
3	5	2	Field, section, flat
4	5	4.6	Storage, airport, convenient
5	6	3.83	Environment, rest,world, answer
6	5	2.4	Folder, label, credit
7	8	4.125	Profile, background, scene, rock
8	5	6.6	Tray, blueray, manual, gripe
9	6	7.83	Alarm, longer, beep
10	5	19.2	Police, ticket, state, ka

4. Results and analysis

The proposed approach is provided 50 datasets, each having hundreds of documents about their specific domain covering various electronic products (Chen, 2014). The reviews are real users data extracted from the Amazon store. In the learning phase, the model processes all the domains and constructs one giant complex network from it. The communities extracted are evaluated for their quality. Communities with weakly associated nodes cannot effectively communicate their impact and are therefore filtered out. Only 84 out of 148 communities were left behind after the *W ttCI*. The threshold value for *W ttCI* is set to 1. This value is experimentally evaluated for the given dataset. Dropping the thresholds adds to computational cost without showing reasonable improvement in results. Table 1 shows properties, i.e., size, and an average degree of a node along with top few words for randomly selected 10 rules.

The empirical log-likelihood (ELL) and mutual information (MI) are calculated to the topics generated. ELL shows the generality of the model as its ability to estimate the unseen data when trained on seen data. Its value should preferably increase as the model makes a better prediction of the unseen data. The ELL maintains its value consistently above −140, even when a number of topics generated were increased from initially 15 to 20, 30 and 40. The MI of the model is maintained above 1. It is also depicted in Figure 1.

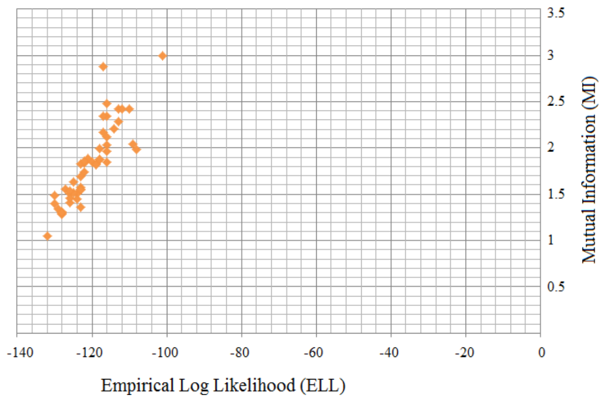

Figure 1. ELL and MI of topics

The proposed approach outperforms the existing approaches of LTM by introducing a more comprehensive mechanism for extracting rules. The test domains are provided to the model following a five-fold cross-validation. All the lifelong learning topic models are provided with 50 domains in their learning phase. Our previously proposed OAMC model does not produce the best results in this situation as it is specifically designed for a streaming sequence of domains. The topic coherence as a measure of the quality of topics produced indicates that the proposed model has average topic coherence of −844 that is better than the topic coherence of OAMC, i.e., −864 and the topic coherence of baseline LDA, i.e., −884.

5. Conclusion

The proposed approach using complex network analysis for learning rules that are consumed by the lifelong learning topic model shows satisfying results. It indicates a positive direction to use complex network analysis for learning rules and benefit from the rich properties of networks and communities. Complex network analysis is highly informative and can be further explored to enhance the learning capabilities of the model. In future, it can lead to generate hierarchical structure of topics within a domain by utilising the structural information held in the community-based rules.

References

Andrzejewski, D. & Zhu, X. (2009). *Latent dirichlet allocation with topic-in-set knowledge*. In Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies NAACL: HLT 2009 Workshop on Semi-Supervised Learning for Natural Language Processing, pp. 43–48, Association for Computational Linguistics.

Aziz, F., Hancock, E. R. & Wilson, R. C. (2016). *Graph entropy from closed walk and cycle functionals*. In Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR), pp. 174–184, Springer.

Blei, D. M. & Jordan, M. I. (2006). Variational inference for dirichlet process mixtures. *Bayesian analysis, 1*(1), 121–144.

Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003) Latent dirichlet allocation. *Journal of Machine Learning Research, 3*, 993–1022.

Bouma, G. (2009). Normalized (pointwise) mutual information in collocation extraction. Proceedings of German Society of Computational Linguistics & Language Technology (GSCL), pp. 31–40.

Chen, Z. & Liu, B. (2014a). *Topic modeling using topics from many domains, lifelong learning and big data*. In Proceedings of the 31st International Conference on Machine Learning (ICML-14), pp. 703–711.

Chen, Z. & Liu, B. (2014b). *Mining topics in documents: standing on the shoulders of big data*. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 1116–1125, ACM.

Chen, Z. & Liu, B. (2016). Lifelong machine learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning, 10*(3), 1–145.

Chen, Z. B. (2014). Chen2014 dataset. Retrieved January 8 2014 from https://www.cs.uic.edu/zchen/downloads/KDD2014-Chen-Dataset.zip

Chen, Z., Mukherjee, A. & Liu, B. (2014). *Aspect extraction with automated prior knowledge learning*. In 52nd Annual Meeting of the Association for Computational Linguistics, pp. 347–358.

Han, L., Escolano, F., Hancock, E. R. & Wilson, R. C. (2012). Graph characterizations from von neumann entropy. *Pattern Recognition Letters, 33*(15), 1958–1967.

Information Resources Management Association (2016). *Big data: concepts, methodologies, tools, and applications*. IGI Global.

Khan, M. T. & Khalid, S. (2016). *Multimodal rule transfer into automatic knowledge based topic models*. In Multi-Topic Conference (INMIC), 2016 19th International, pp. 1–6, IEEE.

Khan, M. T. & Khalid, S. (2016). Sentiment analysis for health care. In *Big Data: Concepts, Methodologies, Tools, and Applications* (pp. 676–689), IGI Global.

Khan, M. T. & Khalid, S. (2017). *Paradigmatic and syntagmatic rule extraction for lifelong machine learning topic models*. In Electrical Engineering and Informatics (ICEEI), 2017 6th International Conference on, pp. 1–6, IEEE.

Khan, M. T., Azam, N., Khalid, S. & Yao, J. (2017). A three-way approach for learning rules in automatic knowledge-based topic models. *International Journal of Approximate Reasoning, 82*, 210–226.

Khan, M. T., Durrani, M., Ali, A., Inayat, I., Khalid, S. & Khan, K. H. (2016). Sentiment analysis and the complex natural language. *Complex Adaptive Systems Modeling, 4*(2), 1–19.

Khan, M. T., Durrani, M., Khalid, S. & Aziz, F. (2016). Lifelong aspect extraction from big data: knowledge engineering. *Complex Adaptive Systems Modeling, 4*(1), 1.

Khan, M. T., Durrani, M., Khan, K. H., Ali, A. & Khalid, S. (2015). Aspect-based sentiment analysis on a large-scale data: Topic models are the preferred solution. *Bahria University Journal of Information & Communication Technology, 8*(2), 22.

Khan, M. T., Yar, S., Khalid, S. & Aziz, F. (2016). *Evolving long-term dependency rules in lifelong learning models*. In Knowledge Engineering and Applications (ICKEA), IEEE International Conference on, pp. 93–97, IEEE.

Mahmoud, H. (2008). *Polya urn models*. CRC press.

Massoulie, L. (2014). *Community detection thresholds and the weak ramanujan property*. In Proceedings of the forty-sixth annual ACM symposium on Theory of computing, pp. 694–703, ACM.

Mimno, D., Wallach, H. M., Talley, E., Leenders, M. & McCallum, A. (2011). *Optimizing semantic coherence in topic models*. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 262–272, Association for Computational Linguistics.

Mowshowitz, A. & Dehmer, M. (2012). Entropy and the complexity of graphs revisited. *Entropy, 14*(3), 559–570.

Pang, B. & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval, 2*(1–2), 1–135.

Passerini, F. & Severini, S. (2008). *The von neumann entropy of networks*. Available at SSRN 1382662.

Peng, S., Yang, A., Cao, L., Yu, S. & Xie, D. (2017). Social influence modeling using information theory in mobile social networks. *Information Sciences, 379*, 146–159.

Ramage, D., Hall, D., Nallapati, R. & Manning, C. D. (2009). Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora. *Proceedings of the Empirical Methods in Natural Language Processing, 1*, 248–256).

Ruvolo, P. & Eaton, E. (2013). Ella: an efficient lifelong learning algorithm. *Proceedings of the 30th International Conference on Machine Learning (ICML-13), 28*, 507–515.

Sadilek, A., Kautz, H. A. & Silenzio, V. (2012). *Modeling spread of disease from social interactions*. In International Conference on Web and Social Media (ICWSM).

Shen, L., Wu, L. & Li, Z. (2016). Topic modelling for object-based classification of vhr satellite images based on multiscale segmentations. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 41*, 359–363.

Xu, M., Yang, R., Harenberg, S. & Samatova, N. F. (2017). *A lifelong learning topic model structured using latent embeddings*. In Semantic Computing (ICSC), 2017 IEEE 11th International Conference on Semantic Computing, pp. 260–261, IEEE.